



## Measuring Depression in Persons With Spinal Cord Injury: A Systematic Review

Claire Z. Kalpakjian, PhD; Charles H. Bombardier, PhD; Katherine Schomer, MA; Pat A. Brown, EdD; Kurt L. Johnson, PhD

Model SCI System, Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan

Received March 26, 2008; accepted June 30, 2008

### Abstract

**Background/Objective:** Depression has been studied extensively among people with spinal cord injury (SCI). However, basic questions persist regarding the reliability and validity of depression measurement in the context of SCI. The objective of this study was to evaluate the state of knowledge of depression measurement in persons with SCI.

**Methods:** English-language peer-reviewed citations from MEDLINE, CINAHL, PsycINFO, ProQuest, Google Scholar, and Web of Science from 1980 to present. Two reviewers screened 377 abstracts on SCI and depression topics to identify 144 containing classifiable psychometric data. All 144 were reviewed by 6 reviewers. Twenty-four studies reporting psychometric data on 7 depression measures in SCI samples were identified, including 7 validity studies.

**Results:** Reliability data were limited to internal consistency and were consistently good to excellent across 19 studies. Validity data were limited to concurrent validity, construct validity, and/or clinical utility in 10 studies. Measures were comparable with respect to internal consistency, factor structure, and clinical utility. Results are limited to peer-reviewed, English literature, and studies were not judged for quality.

**Conclusions:** Greater attention should be paid to the psychometric evaluation of established measures. Although existing evidence may not justify universal screening, we recommend depression screening in clinical practice when patients may be seen by nonpsychology personnel. There is insufficient evidence to recommend one screening measure over another. Therefore, selection of measures will depend on clinician preferences. Psychometric studies are needed to show test-retest reliability, criterion validity, and sensitivity to change to improve depression recognition and treatment.

*J Spinal Cord Med.* Feb 2009;32(1):6–24

**Key Words:** Depression; Screening; Psychometrics; Spinal cord injuries

---

Please address correspondence to Claire Z. Kalpakjian, PhD, Model SCI System, University of Michigan, Department of Physical Medicine and Rehabilitation, 300 N. Ingalls, Room NI 2A09, Ann Arbor, MI 48109; p: 734 763 9374; f: 734 936 5492 (e-mail: [clairez@umich.edu](mailto:clairez@umich.edu)).

The National Institute for Disability and Rehabilitation Research, Office of Special Education and Rehabilitative Services, US Department of Education, Washington, DC, funds the University of Michigan Model Spinal Cord Injury Care System (H133N060032), the University of Washington Northwest Regional Spinal Cord Injury System (H133N060033) and a Disability and Rehabilitation Research Project (H133A060107), and the University of Washington Model Systems Knowledge Translation Center (H133A060070).

© 2009 by the American Paraplegia Society

### INTRODUCTION

Depression has been studied extensively among people with spinal cord injury (SCI) (1), and depression symptoms are estimated to be highly prevalent in this population (2). Depressive symptoms are associated with a myriad of negative outcomes among persons with SCI including lower functional independence (3), more secondary complications (4), poorer community and social integration (5–7), and lower self-appraised health (8). The prominence of depression after SCI is supported by this disorder being the subject of one of the first clinical practice guidelines published by the Paralyzed Veterans of America in 1998 (9). However, the use of depression measures within the SCI population has been criticized on both conceptual and psychometric grounds (1). Given the critical role of depression in the health and well being of persons with SCI, accurate screening for

depression and measurement of symptom severity are crucial if progress is to be made in both research and clinical practice. Thus, examining the state of the science with respect to the measurement of depression in persons with SCI is an important undertaking.

### **DEFINING AND MEASURING DEPRESSION**

The word “depression” is used to describe a variety of states—it can refer to a mood state, a symptom of several disorders, or a syndrome or collection of symptoms that frequently occur together or one of several disorders (10). In the SCI literature, depression generally refers to either a constellation of symptoms or a diagnosis of major depressive disorder (MDD). Major depressive disorder is a clinical diagnosis based primarily on subjective symptoms and exclusion of competing diagnoses. Symptoms are evaluated with respect to intensity, duration, and impact on daily functioning. The Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) (11) defines many disorders including MDD on the basis of the presence of a minimum number of symptoms or features from a list (12). Diagnosis of MDD requires the presence of 5 of 9 psychological and somatic symptoms and must include at least 1 of the 2 essential criteria: depressed mood or loss of interest. The symptoms must be present persistently for at least 2 weeks and cause clinically significant impairment (11). Symptoms of depression are characterized by depressed mood, loss of interest or pleasure (anhedonia), feelings of worthlessness, fatigue, insomnia or hypersomnia, appetite change, weight loss or gain, and suicidality.

Measures of depression include those measuring severity of symptoms (vs a diagnosis) and screening measures that are criterion-referenced. Severity measures may be self-report or observer rated. These scales ask “how much” with respect to depressive symptom severity. When first developed, they were used to evaluate the effectiveness of interventions to reduce symptomatology. One of the earliest self-report depression severity scales is the Beck Depression Inventory (BDI), first developed in 1961 (13). The most widely used observer-rated depression severity measure is the Hamilton Depression Scale (HAM-D), also developed in the 1960s (14). The items comprising depression severity measures do not necessarily correspond to current MDD by DSM criteria but often reflect clinical experience. The BDI and Zung Self-Rated Depression Scale (SDS) (15) were developed in this way. As such, high scores on these measures should not be construed to represent diagnosed MDD according to DSM criteria. Moreover, the item content of some of these measures may be multifactorial, representing anxiety or general psychological distress. Therefore, reviewers such as Elliot and Frank (1) have cautioned researchers not to confuse measures of depressive symptoms or depressive behavior with a diagnosis of MDD. Nevertheless, severity measures have often been used as screening measures with cutoff

scores to indicate “significant” or “clinical” depression. The use of many different measures and different cut-off scores has contributed to rather wide variability in prevalence estimates, as well as the definition of depression and clinical decision making.

With more widespread acceptance of the DSM-based diagnostic system, screening measures were developed for the purpose of clinical and research case-finding. They are criterion-referenced because most are designed to tap into diagnostic criteria, such as the DSM. Although these also cannot be used to diagnose depression, they can indicate a need for further clinical evaluation as they were designed to more closely reflect the symptom content of current diagnostic criteria. For example, the Patient Health Questionnaire-9 (PHQ-9) (16) designates “probable depression” as indicating the need for further evaluation. Psychometric evaluations of these newer measures have focused on criterion-related validity such as sensitivity and specificity. Some can also be used as severity measures.

Diagnostic interviews such as the Structured Clinical Interview for DSM-IV (17) (SCID) or the Composite International Diagnostic Interview (CIDI) are typically considered the “gold standards” against which severity and screening measures are validated. Diagnostic interviews are directly linked to the currently accepted diagnostic nomenclature for MDD. Interviews such as the CIDI are highly structured, permitting nonclinicians to arrive at reliable and valid diagnoses. Others such as the SCID are semistructured and are typically administered by trained clinicians. In some cases, studies will report that DSM criteria were used to diagnose MDD without specifying the manner in which the diagnosis was determined, often vaguely referred to as “psychiatric interview.”

### **Challenges to Measuring Depression in SCI**

One of the primary challenges to measuring depression in the context of SCI is determining whether neurovegetative or somatic symptoms are attributable to the effects of SCI, secondary medical conditions, environmental factors, or to depression itself. This challenge is of course not restricted to SCI, but is present among medical patients, people with disabilities, the elderly, women, children, and adolescents, culturally diverse groups, prison populations, and the poor (18). Williams et al (19) proposed 2 primary approaches to handling the effect of physical illness on the diagnosis of depression, and these provide some framework for considering this review of depression measurement. The inclusive approach counts depressive symptoms toward the diagnosis of depression irrespective of whether the symptom is judged to be caused by medical and psychologic causes. The etiologic approach reflects the DSM’s criteria that count symptoms toward a diagnosis unless the symptom is clearly accounted for by a medical condition.

Somatic symptoms have been a key feature of depression in various nosologic systems including the DSM and as far back as the time of Hippocrates (20). Gastrointestinal symptoms, sleep disturbance, headaches, appetite changes, fatigue, and aches and pains of a diffuse nature are common features of depression. However, these same symptoms are also common across many other medical conditions, including SCI, making distinguishing depression from primary or secondary conditions or medical comorbidities particularly challenging.

The presentation of somatic vs psychological symptoms by patients has been studied extensively in the primary care literature. Under-recognition of depression in those presenting with primarily or exclusively somatic symptoms may be the single most common reason why psychiatric illnesses go undetected (20). Studies have suggested that the vast majority of patients with depression will complain of somatic symptoms rather than psychologic (21,22). There also seems to be a dose-response relationship between somatic complaints and depression, with higher likelihood of depression linked to greater physical complaints and poorer response to treatment (23). It is also well known that pain and depression co-occur at high rates, and it is postulated that they share biological pathways and neurotransmitters (20,24), with pain shown to be a primary complaint in early stages of depression and a risk factor for poor treatment response (20,21).

### **Need for Evaluating Measurement Tools**

The need for valid and reliable tools to measure depression in the SCI population—or any population for that matter—cannot be overstated. In 2 influential reviews, Elliott and Frank (1) and Frank et al (25), as well as others (26), have observed that studies of depression in people with SCI have been fraught with inadequately defined constructs, a diversity of measures purportedly measuring depression and little attention paid to the broader literature on depression, including widely accepted diagnostic criteria. Meaningful and accurate estimates of depression incidence after injury and its prevalence in the SCI population are predicated on the quality of the tools we use to measure depression. Measurement is inextricably bound to our ability to evaluate the effectiveness of our interventions, to make compelling arguments for the utilization of resources to treat depression, and to develop and implement screening and treatment programs. More broadly, because research builds on previous work, the use of measures that have not been validated for the population of interest creates an unstable foundation on which treatment practices or new lines of research may be based. Because of the iterative nature of science—replication and building on what has come before—the need to validate depression measurement tools in this population is critical.

## **METHODS**

The purpose of this paper is to systematically review the peer-reviewed, published literature on psychometric characteristics of depression measures when used in SCI samples. Based on this review, we make recommendations to researchers and clinicians for future work in this area given the importance of accurate depression diagnosis, assessment, and measurement of those with SCI. While there are established methods for conducting systematic reviews for interventions, such as the Cochrane Review for randomized controlled trials, there are no established methods or guidelines for conducting a systematic review of measurement tools. In developing the methods for this review, we considered these established guidelines, but modified them for purposes of conducting our review.

### **Criteria for Considering Studies for Review**

The following criteria were used to select studies of depression in the SCI population in the published literature for this systematic review: (a) a study of depression in persons with SCI; (b) published since 1980; (c) written in English; (d) conducted in adults older than 18 years of age; (e) has a study population that includes persons with SCI (but need not exclusively be SCI); and (f) is peer reviewed. Any study design (ie, randomized controlled, case control, cohort, and case series), review papers, and meta-analyses were included in the initial search.

### **Search Methods for the Identification of Studies**

Separate searches of the published literature were conducted in MEDLINE, CINAHL, PsycINFO, ProQuest, Google Scholar, and the Web of Science. Specific search terms used for each database included (a) depression, (b) major depression, (c) major depressive disorder, and (d) major depressive episode and SCI. In addition, a list of commonly used depression measures was included in search terms to target searches further. Comprehensive mood scales that include depression subscales also were included. Specifically, the following scales were included in the database search: Center for Epidemiologic Studies Depression (CES-D); Beck Depression Inventory (BDI); Patient Health Questionnaire (PHQ-9); Zung Self Rating Scale (SRS); the Hamilton Depression Scale (HDS); Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID); SCL-90; Hospital Anxiety and Depression Scale (HADS); and the Brief Symptom Inventory (BSI). A comprehensive list of search terms and measures by database is included in the Appendix.

Abstracts from the initial database search were reviewed by 2 reviewers who were either graduate research assistants or research staff at the University of Washington Model Systems Knowledge Translation Center (MSKTC). Abstracts were reviewed to identify validation studies of depression measures. Reviewer discrepancies were resolved by consensus of the review-

**Table 1.** Levels of Study Classification

Level	Description
1	Psychometric evaluation of a depression measure is the primary aim of the study; may have secondary aims like describing depression in the sample, but the primary aim of validation or other psychometric characteristic is the focus.
2	Depression is the primary outcome; may have other secondary or tertiary outcomes of interest, but depression is the primary focus. Many times outcome is tested in regression analysis or group comparisons on depression. Depression can also be examined across time with time as the predictor.
3	Depression is a secondary outcome and not the primary focus of the study. May be among multiple outcomes of equal interest.
4	Depression is a correlate, predictor or mediator. Not a primary or secondary outcome. Predictor is mostly identified in regression analysis. As a correlate it is found mostly in bivariate correlation tables; sometimes depression is characterized as an outcome but presented only in bivariate correlations.
5	The study estimated depression prevalence in an SCI sample. Few studies exclusively focus on prevalence without including it as an outcome or predictor; in those cases studies were classified in levels 2, 3 or 4.

ers and MSKTC research staff. Author and journal names were not masked from the reviewers. Only 7 validation studies were identified in the review of abstracts. Thus, an alternative strategy was developed to widen the search for studies examining any of various psychometric properties of depression measures reported in studies of persons with SCI. To organize the studies identified in the initial database search, 5 “levels” were created to categorize and identify studies reporting psychometric characteristics of depression measures, not only validation studies. Descriptions for each of the levels are given in Table 1. When reviewers were unable to assign a classification level based on the title or abstract, the full article was reviewed for classification assignment. If a study did not meet any of the level criteria, it was excluded from further review.

After studies were classified into the levels, 6 reviewers extracted psychometric data by reviewing the manuscript in full. Because articles classified as level 1 are the most important and contain the most measurement validation data, data were extracted by 2 reviewers and compared for consistency. Differences were reconciled by consensus. One third of the remaining articles in levels 2 through 5 were randomly selected and reviewed separately by a second reviewer for consistency. The second reviewer rated the first data extraction of the

article on a scale of 1 to 4, where 1 indicated significant change was needed and 4 indicated little to no changes. Discrepancies requiring significant changes were only observed by one of the randomly selected second reviews. Data were directly entered into a Microsoft ACCESS database specifically designed for article data extraction, thus reducing data entry errors from the use of paper forms or worksheets. In the end, only studies in which psychometric data on measures of depression and SCI were included in the final review reported here.

### Psychometric Data Extracted From Studies

Based on our review of previous work evaluating psychometric characteristics of depression measurement tools in general (27) and other work evaluating measurement tools (28–30), a set of criteria was created to evaluate each of the measurement tools in this review. The primary purpose for these criteria was to find evidence that supports the overall reliability and validity of a depression measure when used with persons with SCI. Other criteria examined the administration of the measure. The following provides a brief description of the psychometric criteria that were the target of data extraction by reviewers.

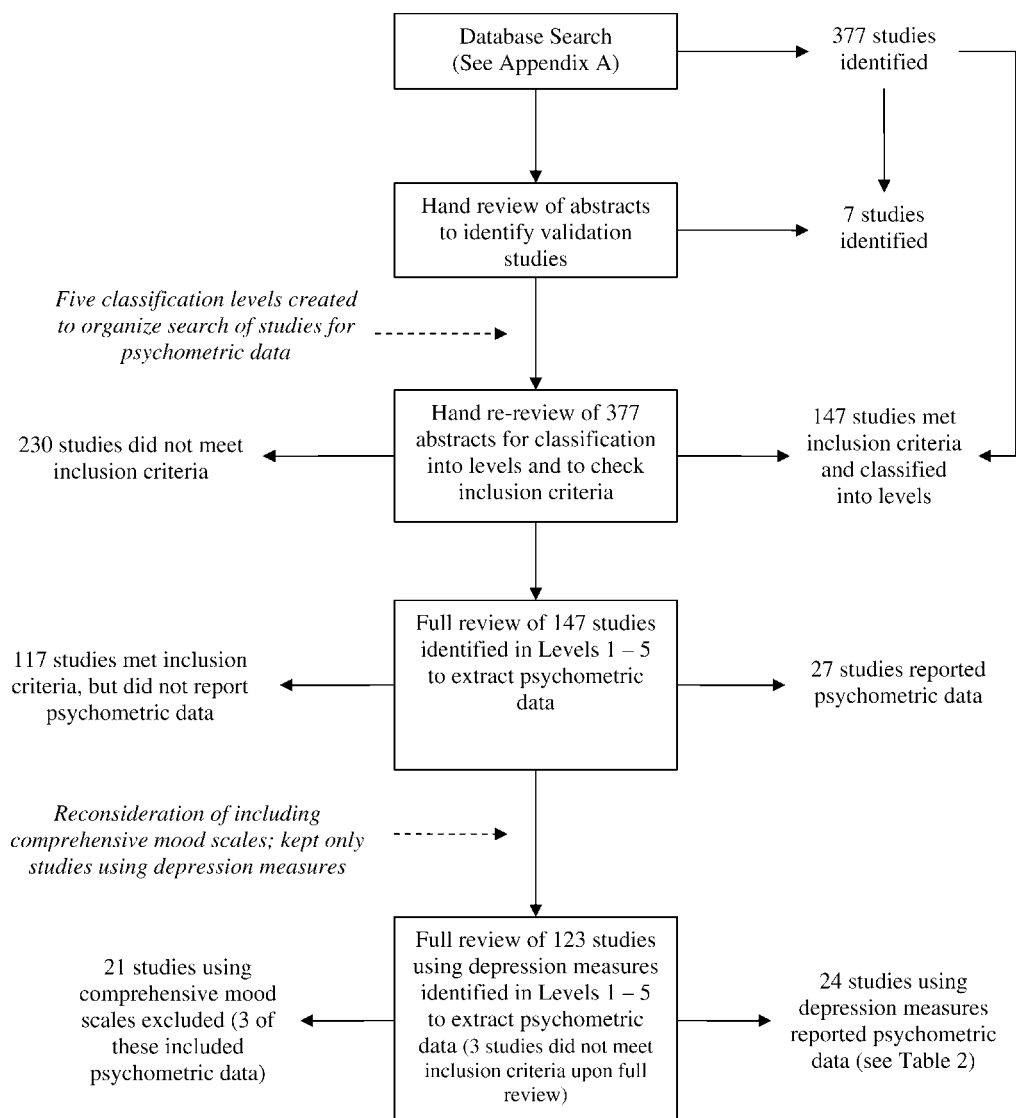
### Measure Administration

In general, the administration of self-report depression measures is brief (5–10 minutes) and typically uses paper/pencil formats or reading items aloud to respondents. Structured or semistructured interviews are more time and labor intensive, requiring a trained evaluator, but ordinarily do not require special accommodations for administration. Measures were examined for the following properties: other languages, disability adaptation, time to administer, administration burden, and other administration comments.

### Reliability

Broadly speaking, reliability refers to the consistency of measurement. All tests have error; reliability focuses on nonsystematic or random errors. When random errors are minimal, scores on a test can be expected to be more consistent or stable across administrations. There is no way to directly observe or calculate a person’s true score, so a variety of methods are used to estimate reliability of a test. Four types of reliability were the focus of this review: internal consistency, test-retest, alternate forms, and ceiling/floor effects. Internal consistency (eg, Cronbach  $\alpha$ ) is interpreted as the mean of all possible split-half coefficients that are possible for a given measure. The  $\alpha$  coefficient is most applicable when a test measures a single construct, such as depression, and is less useful for multidimensional measures. In the test-retest method, reliability is estimated as the Pearson product-moment correlation coefficient between 2 administrations (time between measurements will vary) of the same measure with the same individuals. In the alternate forms method,





**Figure 1.** Study selection.

reliability is estimated by the Pearson product-moment correlation coefficient of 2 different forms of a measure, usually administered at the same time.

### Validity

Broadly speaking, validity refers to the extent that a measure actually measures what it purports to measure. There is no single index of validity; rather validity is estimated by reviewing the quality of several indicators of validity. Six domains of validity were the focus of this review: content validity, scale dimensionality, convergent and discriminant validity, predictive validity, and clinical utility indices.

Content validity refers to the degree to which the content reflected in the items in the instrument adequately sample the content of the construct. For example, depression-screening measures base their content specifically on diagnostic criteria of the DSM-III or DSM-IV. Construct validity refers to the degree to

which the measure represents a well-defined underlying theoretical construct. Concurrent validity refers to the degree to which a measure correlates with another validated measure in the predicted direction; concurrent validity with a gold standard such as the SCID is the cornerstone of validation for depression measures. A depression measure would be expected to converge with scores on other measures of depression and diverge from measures of optimism, for example. The latter is also known as discriminant validity. Predictive validity refers to the degree to which scores on an instrument are associated with outcomes in the future.

Clinical utility reflects the accuracy with which a screening measure can identify cases. Cases are individuals with a given diagnosis, such as DSM-IV–defined MDD. The sensitivity of a measure refers to how well a screening instrument detects a target disorder in a person who actually has the disorder or problem. Specificity refers to the ability of a screening measure to successfully

**Table 2.** Frequency of Depression Measures Represented in Levels 1–5 After Reclassification (n = 123)\*

Measure	Level 1	Level 2	Level 3	Level 4	Level 5	Total
Beck Depression Inventory (BDI)	1	6	20	13	4	42
Center for Epidemiological Studies, Depression (CES-D)	2	4	20	11	0	37
Hamilton Depression Rating Scale (HAM-D)	0	3	0	3	0	6
Inventory to Diagnose Depression (IDD)	2	0	6	3	0	11
Patient Health Questionnaire (PHQ-9)	1	2	0	2	0	5
Older Adult Health and Mood Questionnaire (OAHMQ)	0	4	4	3	0	11
Structured Clinical Interview for the DSM (SCID)	0	0	0	0	1	1
Zung Self-Rating Depression Scale (SDS)	1	0	4	2	1	8
Total number studies by level	8	19	54	35	6	123

\*The 21 studies that used comprehensive mood scales were excluded.

identify those people without the target disorder or problem. Receiver operating characteristics (ROC) curves are used to identify optimal models for diagnostic decision making; ROC curves are a graphical plot of the sensitivity and 1 – specificity. The area under the curve (AUC) refers to the accuracy of a diagnostic test; AUC values >0.80 indicate good-to-excellent diagnostic accuracy.

Likelihood ratios are another way to interpret the values obtained for sensitivity and specificity. A likelihood ratio provides the odds that a given screening instrument would be found to be positive in a person with, as opposed to without, the target problem or disorder. Positive likelihood ratios >10 and negative likelihood ratios <0.10 indicate good discriminability. The positive predictive value (PPV) of a measure refers to the post-test probability of actually having the target disorder or problem for a positive screen, and the negative predictive value (NPV) of a measure refers to the post-test probability of not having the target disorder or problem for a negative screen.

## RESULTS

### Selection of Studies

A total of 377 studies were selected from the initial database search. Of these, only 7 were validation studies. As described above, selection criteria were revised to include reporting of any psychometric data and classification by 5 levels. Of the 377 studies identified in the database search, 147 were classified (by abstracts) into the 5 levels, and 230 were excluded from further review. These 230 studies were excluded because they either could not be classified or did not meet full criteria for inclusion in the review. For example, many did not actually address depression, did not use standardized depression measures, or did not use SCI samples or report separate data for an SCI subsample.

A full review of the 147 studies identified 27 that reported psychometric data. The other 117 studies met inclusion criteria but did not report any psychometric data, and another 3 were identified as not actually meeting inclusion criteria when fully reviewed. Further-

more, the inclusion of studies using depression subscales from comprehensive mood scales (ie, SCL-90) or depression and anxiety scales (ie, HADS) was reconsidered after the psychometric data were extracted. It was decided that only measures designed to measure depression or depressive symptoms would be included in this review. Thus, only 24 studies using depression measures reported psychometric data and are the focus of this review. A flow chart of the study selection is shown in Figure 1.

### Characterizing the SCI Literature Using Depression Measures

After the initial classification of studies into the 5 levels was implemented and psychometric data were extracted, a review of the initial assignment of studies (which was done only using the abstract information) was reconsidered as a way to characterize the population of studies using these depression measures in the extant SCI literature. This secondary review (and reclassification in some cases) of studies should be considered as an addendum to the systematic review because the classification system was originally intended only as a way to organize the studies. However, after the psychometric data were extracted, it was found to be a useful method of summarizing the use of the different depression scales in the extant SCI literature. The first author (C.Z.K.) rereviewed the design of all 123 studies using depression measures (excluding the 21 studies that used comprehensive mood scales) to confirm level assignment. A summary of study classification by depression measure is given in Table 2.

As we described, we initially included both comprehensive mood scales and depression measures in the database search. The BDI, CES-D, SDS, SCID, HAM-D, and PHQ-9 were identified a priori for the database search, and, with the exception of the HAM-D, all were used in studies reporting psychometric data. Our search strategy allowed us to also identify 2 measures that were not considered a priori, but for which psychometric data for SCI samples were reported: the Older Adult Mood & Health Questionnaire (OAHMQ) and Inventory to Diag-

**Table 3.** Summary of Depression Measure Characteristics

Measure	Year Published	Type	Theoretical Orientation	Content	Scoring
Beck Depression Inventory (BDI) (13)	1961; 1971 (BDI-1A); 1996 (BDI-II)	Severity	Two-dimensional; psychological/cognitive and physical/somatic	21 items for BDI; 13 and 7 items for BDI-II.	Items rated on 4-point scales ranging from least to greatest severity. Cut score of 14 for depressed vs not depressed.
Zung Self-Rating Depression Scale (SDS) (15)	1965	Severity	Multidimensional; pervasive negative affect, physiological, and psychological	20 items; 10 negatively and 10 positive worded items.	Items rated on 4-point Likert scales ranging from none of the time to most of the time. No cut score; higher score is worse severity.
Center for Epidemiological Studies, Depression (CES-D) (82)	1977	Severity	Designed to estimate depressive symptoms in general population; emphasis on depressed mood	20 items; 4 subscales (depressive affect, somatic, positive affect and interpersonal relations)	Items rated on 4-point Likert scales ranging from rarely or none of the time to most or all of the time. Cut score of 16 recommended for indicating possible depression in previous week.
Older Adult Health and Mood Questionnaire (OAHMQ) (60)	1995	Severity	Related to DSM-III criteria; took into account age differences and to distinguish depression vs normal aging	22 items	Items rated as true or false with scores ranging from 1 to 22. Cut score of $\geq 11$ for probable major depressive episode.
Inventory to Diagnose Depression (IDD) (83)	1986	Screening	Based on DSM-III criteria for presence/absence as well as duration	22 items	Each item is comprised of a group of five statements that reflects increasing severity of that symptom. When endorsed, the item is then rated for duration of $\leq 2$ weeks.
Patient Health Questionnaire – 9 (PHQ-9) (84)	1990s	Screening	Based on DSM-IV criteria	10 items; 10th item assesses difficulty symptoms have caused in daily life	Items rated on 4-point Likert scales ranging from not at all to nearly every day. Two scoring methods for probable MDD: (a) depressed mood and/or anhedonia plus 4 symptoms for more than half the days or (b) total score $\geq 10$ . Decline of 5 points considered clinically significant response.
Structured Clinical Interview for DSM-IV (SCID) (85)	1990	Diagnostic interview	Is the DSM criteria	10 modules of diagnostic criteria DSM disorders	Presence or absence of criteria is scored on a 1 (false) to 3 (true) scale; 2 is subthreshold.

**Table 4.** Overview of Studies Reporting Psychometric Data (N = 24)

Depression Measure	Author and Article Citation	Country	Sample Size	Sample Description	Psychometric Data Reported
BDI	Radnitz et al 1997 (33)	United States	124	Veterans (inpatient and outpatient) from VA (all males)	IC and CCV
BDI	Ferrington et al 1986 (34)	United States	104	Adult inpatients (all males) >5 months post traumatic SCI	IC
CES-D	Kupniratsaikul et al 2002 (79)	Thailand	83	Not specified with respect to inpatient/outpatient status	CCV
CES-D	McColl et al 1992 (45)	Canada	120	Adults recently discharged (<1 month) from inpatient rehabilitation (new SCI)	IC and CSV
CES-D	Crisp et al 1992 (43)	United States	60	Adults >5 years postinjury (long-term follow-up study)	IC
CES-D	Dorsett et al 2004 (42)	Australia	46	>16 years old with acute traumatic injury; assessed at baseline (discharge) and 3 years postdischarge	IC
CES-D	Latimer et al 2004 (37)	Canada	21	Adults with chronic SCI participating in randomized controlled trial to test exercise and mood	IC
CES-D	Shnek et al 1997 (46)	United States	80	Adults with chronic SCI	IC
CES-D	Ginis et al 2003 (39)	Canada	34	Adults with chronic SCI (>1 year postinjury)	IC
CES-D	Coyle et al 1993 (41)	United States	48	>16 years old, had ≥2 hospitalizations in previous years caused by health problems	IC
CES-D	Hicks et al 2003 (38)	Canada	34	Adults with traumatic SCI (>1 year postinjury)	IC
CES-D	Latimer et al 2005 (36)	Canada	33	Adults with chronic SCI (did not specify time since injury)	IC
CES-D	Rintala et al 2005 (44)	United States	165	Veterans with chronic SCI (≥6 months postinjury)	IC
CES-D	Decker and Schulz 1985 (86)	United States	100	Adults ≥40 years old with chronic SCI (>5 years postinjury)	IC
SDS	Tate et al 1993 (59)	United States	162/30	Adults with chronic SCI/inpatients with new traumatic SCI	IC, CSV, and CCV
PHQ-9	Bombardier et al 2004 (2)	United States	849	Adults with traumatic SCI at 1-year follow-up postinjury	IC and CSV*
PHQ-9	Kalpakjian and Albright 2006 (54)	United States	1,168	Adults with chronic SCI (case-matched pairs of men and women)	IC
OAHMQ	Krause et al 2000 (49)	United States	1,391	Adults with chronic (>1 year) traumatic SCI	IC and CSV
OAHMQ	Kemp et al 1999 (51)	United States	171	Adults with chronic (>2 years) traumatic SCI (mixed ethnic backgrounds)	CSV
OAHMQ	Krause et al 1999 (50)	United States	97	Adults with chronic (>1 year) traumatic SCI with American Indian ethnic background	IC
OAHMQ	Kemp et al 2004 (48)	United States	43	Adults living in the community	IC
IDD	Frank et al 1992 (52)	United States	134	Adult inpatient and outpatients	CSV
IDD	Clay et al 1995 (53)	United States	133	Adults with chronic tetraplegia and paraplegia	CCV
SCID	Radnitz et al 1996 (31)	United States	125	Veterans with SCI (50 control veterans who experienced comparable trauma but no SCI)	Test-retest

\*Evaluated PHQ-9 items against PHQ-9 score vs other diagnostic criteria.

IC, internal consistency (see text for values); CCV, concurrent validity; CSV, construct validity using factor analysis (see Table 5 for details); BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; SDS, Zung Self-Rated Depression Scale; OAHMQ, Older Adult Mood and Health Questionnaire; IDD, Inventory to Diagnose Depression; SCID, Structured Clinical Interview for the DSM.



**Table 5.** Sensitivity, Specificity, and Other Clinical Utility Indices Reported Across Reviewed Studies (N = 4)

Article Citation	Diagnostic Instrument	Depression Measure and Cut Score	Sensitivity	Specificity	PPV	NPV	AUC or Efficiency
Radnitz et al (33)	MDD criteria (DSM-III-R)	BDI, cut score 18	83.3%	90.8%	—	—	—
Radnitz et al (33)	MDD criteria (DSM-III-R)	BDI, cut score 27	100%	50.0%	—	—	—
Tate et al (35)	Clinical interview, DSM-III-R criteria	SDS, cut score 55	86.0%	67.0%	—	—	—
Kuptniratsaikul et al (79)	Psychiatric interview (criteria not specified)	CES-D, cut score 19	80.0%	69.8%	45.7	91.7	0.826
Clay et al (53)	MDD criteria (DSM-III)	IDD, Symptom absence or presence* 1. Depressed mood 2. Loss interest/pleasure 3. Energy 4. Psychomotor 5. Guilt 6. Suicide 7. Concentration 8. Appetite 9. Sleep	<div><div>P</div><div>T</div><div>0.71</div><div>0.43</div></div>	<div><div>P</div><div>T</div><div>0.94</div><div>0.84</div></div>	<div><div>P</div><div>T</div><div>0.56</div><div>0.30</div></div>	<div><div>P</div><div>T</div><div>0.97</div><div>0.90</div></div>	<div><div>P</div><div>T</div><div>Y</div><div>N</div></div>

\*Values shown are paraplegia (P) vs tetraplegia (T).

BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; SDS, Zung Self-Rated Depression Scale; IDD, Inventory to Diagnose Depression.

nose Depression (IDD). We also identified the HADS, SCL-90, and BSI in studies reporting psychometric data (3 studies in total), but as we have described, we decided to focus only on those measures specifically designed to measure depression. Excluding these comprehensive scales, 8 depression measures are included in these results. A summary of the characteristics of these depression measures is provided in Table 3, and a summary of the 24 studies and the measures they used that were included in the review are given in Table 4.

In general, evaluation of reliability across all measures was limited to internal consistency with the exception of 1 study using the SCID (31). Of the 24 studies, 19 reported internal consistency (Cronbach  $\alpha$ ) for the depression measures. Of these 19, 14 only reported internal consistency and no other reliability or validity data. Validity fared somewhat better than reliability but was still limited to concurrent validity, clinical utility, and scale dimensionality. Concurrent validity against diagnostic criteria was examined in 5 studies and is summarized in Table 5. Five studies examined scale dimensionality using factor analysis and are summarized in Table 6. No studies examined content validity, convergent validity with other depression measures, or predictive validity of the depression measures. With respect to administration of the depression measures, when accommodations were reported, they were primarily reading items aloud to respondents.

In the next section, we present the results for each of the 8 measures (grouped by severity measures, screening measures, and diagnostic interviews, respectively) in detail. For each, we first describe the measure and then the results of the review for that measure. We also present the number of studies of the 144 reviewed that included the measure and then the number of those who reported psychometric data.

### Depression Severity Measures

**Beck Depression Inventory. Summary.** The BDI was used in 44 studies (30.5% of the 144 reviewed studies), and of these, only 2 reported psychometric data; a third (32) inferred psychometric data but did not report statistical support for their conclusions (we have included that study here).

**Internal Consistency.** Internal consistency coefficients for the 2 studies were excellent (0.89) (33) and (0.84) (34).

**Concurrent Validity.** Radnitz et al (33) performed a discriminant function analysis among 124 veterans with SCI and examined concurrent validity with the SCID (DSM III-R patient edition) to determine how well the BDI items discriminated between depressed and nondepressed subjects (as determined by the SCID). The proportion of variance accounted for by the BDI items was 59%. Items with loadings <0.30 were considered poor discriminators of depressed subjects; this involved somatic preoccupation (item 20), weight

**Table 6.** Summary of Construct Validity (Factor Analyses) Studies of Depression Measures Reviewed (N = 5)

Depression Measure	Factor Analytic Method (exploratory type; rotation; loading threshold; no. factors; % total variance)	First Factor, % Variance (Representative Items)	Second Factor, % Variance (Representative Items)	Third Factor, % Variance (Representative Items)	Fourth Factor, % Variance (Representative Items)
SDS (35)	Principal components analysis; varimax rotation; loadings >0.50; 2 factors; 35% total variance	Affective; DNR% <i>hopeful, life is full, enjoy things</i>	Psycho/somatic, DNR% <i>fatigue, sleep disturbance, crying</i>		
CES-D (45)	Principal components analysis; Lisrel; DNR; 1 factor; 39.6% total variance	Did not label, 39.6% <i>all items</i>			
OAHRMQ (51)	Principal axis analysis; varimax rotation; loadings >0.40; 5 factors; 58.4% total variance	Anhedonia and helplessness, 34.7% <i>little pleasure from anything</i>	Affective, DNR% <i>crying, sleep disturbance, sadness</i>	Energy, DNR% <i>trouble getting started</i>	Slowing down, DNR% and detachment and disinterest, DNR%
OAHRMQ (49)	Principal components analysis; varimax rotation; loadings >0.40; 3 factors; 45% total variance	Evaluative, 31% <i>negative evaluation of life, hopelessness</i>	Affective, DNR% <i>sadness, tearfulness</i>	Behavioral change, DNR% <i>declined activity</i>	
IDD (52)	Un-weighted least squares; Varimax rotation; >0.35; 4 factors; unclear % total variance for SCI group	Core aspects of depression, DNR% <i>low mood, guilt, suicidal ideation, hopelessness</i>	Reports other factors as representing aspects of the SCI itself, but does not specify		

Name of factors as given by author(s) in corresponding study.

DNR, did not report; DNR%, did not report % variance accounted for by the individual factor; SDS, Zung Self-Rated Depression Scale; OAHRMQ, Older Adult Mood and Health Questionnaire; IDD, Inventory to Diagnose Depression.

loss (item 19), body image change (item 14), and work difficulty (item 15).

Various cut scores of the BDI also were examined in this study (33) to determine optimal thresholds, depending on whether there was a need to identify those who were not depressed or those who were. At a cut score of 18, the BDI had a sensitivity of 83.3% and specificity of 90.8%, and 90.0% of all subjects were correctly classified compared to a diagnosis of MDD using the SCID. Alternatively, a cut score of 27 had a sensitivity of 100% but a specificity of only 50%; 95.2% of all subjects were correctly classified.

Judd et al (32) also examined the utility of the cut score of 14 among a sample of 71 newly traumatically injured patients on an inpatient unit. Subjects were evaluated weekly (the total duration of evaluation was not stated) using the BDI. Subjects scoring over 14 also underwent a clinical evaluation to confirm depressed mood (the nature of the clinical or “psychiatric” assessment is not described). The majority (62%) of the sample had consistently low scores of <14 on the BDI; 18% had what was termed “isolated” scores of >14 and were deemed as having “understandable dysphoria” on further clinical examination; and 20% had consistently elevated BDI scores of >14 and were confirmed depressed by clinical interview. Although statistical evidence was not presented, the authors conclude that items most likely to distinguish those who were and were not depressed were sadness, somatic preoccupation, anorexia, pessimism, guilt, irritability, and suicidal ideas. They further note that some items such as weight loss were endorsed by all subjects and were not of discriminative value.

**Zung SDS. Summary.** The SDS was used in 8 studies (5.5% of the 144 studies reviewed), with only 1 reporting psychometric data (however, this study used the SDS in 2 samples and report psychometric data for both).

**Internal Consistency.** Internal consistency was reported in 1 study reviewed and was excellent (0.8135).

**Scale Dimensionality.** Tate and colleagues (35) examined the dimensionality of the SDS (principal components analysis; Varimax rotation; loadings >0.50 retained) in a sample of 162 outpatients with SCI and found a two-dimensional structure for the SDS. The affective dimension was made up of items such as feeling useful, life is full, enjoying things, hopeful about the future, enjoying sex, and easy to do things. The psychosomatic dimension was made up of items such as being tired for no reason, trouble sleeping, irritability, crying spells, restlessness, and constipation. Together, the 2 dimensions accounted for 35% of the total variance of the SDS total score. It is interesting to note that these 2 dimensions also represent positively (affective) and negatively worded items (psycho/somatic).

**Concurrent Validity.** In another sample of the study above (35), 30 inpatient subjects were assessed for depression by a physician and either a psychologist or

social worker. Clinicians were instructed to rate the respondent as depressed or not depressed using DSM-III-R criteria. A subject was deemed depressed if both clinicians rated them as depressed; inter-rater reliability between clinicians was high ( $\kappa = 0.63$ , 84% agreement). At a cut score of >55 (considered to be “depressed”), SDS sensitivity was high at 86% and specificity was modest at 67%.

**Center for Epidemiological Studies Depression Scale. Summary.** The CES-D was the second most widely used scale, represented in 37 studies (25% of the 147 studies review), and of these, 12 reported psychometric data.

**Internal Consistency.** Of the 12 studies reporting psychometric data, 11 reported only internal consistency, which ranged from good to excellent (>0.7036–38; 0.8339, 40; 0.8641; 0.86–0.9442; 0.8843; 0.9044; 0.9145; and 0.9246).

**Scale Dimensionality.** McColl et al (45) examined the dimensions of the CES-D in a sample of 120 adults with SCI who had completed acute and outpatient rehabilitation. Principal components analysis was conducted on the items to determine whether a unidimensional model was empirically reasonable. After confirming that a one-dimensional model could be submitted for further analysis (based on eigenvalue differences), model fit was examined (using LISREL) by root mean-squared residuals (RMSR) <0.150, coefficient of determination (CD) >0.90, and adjusted goodness of fit (AGF) >0.80. Results indicated that a one-factor model of the CES-D in this sample had a satisfactory fit (RMSR = 0.103, CD = 0.951, and AGF = 0.942).

**Concurrent Validity.** Kuptniratsaikul and Pekuman (47) used diagnostic criteria for MDD (DSM-IV) from a “psychiatric interview” to examine the validity of a cut score of 19 in a sample of 83 Thai patients with SCI. A cut score of 19 (vs the recommended score of 16 in a general population) had 80% sensitivity and 69.8% specificity. In that same study, the PPV for a cut score of 19 was 45.7%, and the NPV was 91.7%; the AUC was 0.83, indicating good diagnostic accuracy.

**Older Adult Health and Mood Questionnaire. Summary.** The OAHMQ was used in 11 studies (7.6% of the 144 studies reviewed), and of these, 4 reported psychometric data.

**Internal Consistency.** Three studies reported excellent internal consistency (0.8648, 0.9049, and 0.9150).

**Scale Dimensionality.** Kemp and Krause (51) examined the factor structure (principal axis analysis; Varimax rotation; loadings >0.40 were retained) of the OAHMQ in 171 adults with SCI. Five factors emerged accounting for 58.4% of the variance; the first factor accounted for 34.7% of the variance and reflected anhedonia and a sense of helplessness. The second factor reflected affective symptoms such as crying and sleep loss; the third factor was related to energy; and the fourth reflected a slowing down of activities and regrets about the past. The last reflected detachment and

disinterest in life. Cronbach  $\alpha$  coefficients for the factors ranged from 0.51 (fifth) to 0.84 (first).

Krause et al (49) conducted a factor analysis (principal components analysis; Varimax rotation; loadings over :0.40 were retained) of the OAHMQ in a sample of 1,391 adults with SCI. Three factors emerged from the analysis, accounting for 45% of the total variance; 31% of this variance was associated with the first factor. The first factor was labeled as evaluative and reflected negative evaluation of life and hopelessness about the future ( $\alpha = 0.81$ ); the second was affective reflecting sadness and tearfulness ( $\alpha = 0.82$ ); and the third was behavioral change toward fewer activities ( $\alpha = 0.61$ ).

## Depression Screening Measures

*Inventory to Diagnose Depression. Summary.* The IDD was used in 11 studies (7.6% of the 144 studies reviewed), 2 of which reported psychometric data.

*Reliability.* No studies examined any form of reliability of the IDD.

*Scale Dimensionality.* Frank et al (52) examined the dimensionality (unweighted least squares; Varimax rotation; loadings :0.35 retained) of the IDD in 134 SCI outpatients (along with rheumatoid arthritis, student, and community samples; factor analysis was performed separately for each group). A 4-factor structure emerged with a primary dimension (factor 1) related to affective and cognitive aspects of depression and reflecting the “core aspects” of major depression. Items loading on this factor included low mood, guilt, worthlessness, suicidal ideation, anxiety, and hopelessness. For SCI, the second factor reflected common SCI sequelae such as decreased libido, sleep, appetite, energy, and psychomotor retardation. (The authors do not specify clearly the other factors for any of the groups.)

*Concurrent Validity.* Clay et al (53) examined the accuracy of depression diagnosis in persons with SCI and in particular examined the efficiency of IDD items against DSM-III-R criteria for depression. Specifically, they examined the base rate, sensitivity, and specificity of the IDD items against diagnostic criteria by level of injury (ie, paraplegia,  $N = 80$  vs tetraplegia,  $N = 53$ ). They were most interested in using a Bayesian approach to examine the likelihood of the presence or absence of a depression diagnosis given the presence or absence of individual symptoms by using positive and negative predictive power. The efficiency of the item (representing the 9 diagnostic symptoms) was determined by the ratio of the base rate to 1—base rate must exceed the ratio of false positive ( $1 - \text{specificity}$ ) to true positive (sensitivity). This analysis found that, for both persons with paraplegia and tetraplegia, lack of interest or pleasure, psychomotor disturbance, concentration difficulties, appetite change, and sleep disturbances met efficiency criteria. Dysphoric mood and suicidal ideation efficiently predicted depression for those with paraplegia, but not

tetraplegia; reduced energy level was efficient for predicting depression in those with tetraplegia but not paraplegia. For those with paraplegia, lack of interest or pleasure was the best diagnostic indicator of depression; for those with tetraplegia, inability to concentrate was the best predictor of depression. See Table 4 for details on the values for each of the 9 diagnostic symptoms.

*Patient Health Questionnaire-9. Summary.* The PHQ-9 was used in 5 studies (3.5% of the 144 studies reviewed), and of these, 2 reported psychometric data.

*Internal Consistency.* In both studies reported psychometric data, internal consistency was excellent (0.872 and 0.8654); neither examined any other form of reliability.

*Concurrent Validity.* Concurrent validity of the PHQ-9 with a diagnostic interview or the SCID has not yet been reported. Bombardier et al (2) did examine individual items as predictors of probable MDD as indicated by the PHQ-9 score (using the original scoring scheme). Using a threshold of  $>80\%$  sensitivity, items of depressed mood, disturbed sleep, decreased energy, anhedonia, and feelings of failure were indicators of probable MDD. All symptoms had relatively low PPV. Alternatively, the NPV of all items was high, indicating a high probability of not having probable MDD when none of the items are reported. Positive likelihood ratios ranged from 5 to 1 for sleep disturbance to 18 to 1 for psychomotor changes; negative likelihood ratios were less robust.

## Diagnostic Interviews

*Structured Clinical Interview for DSM-IV. Summary.* The SCID was identified in 1 of the 144 studies reviewed.

*Reliability.* Radnitz et al (31) used the SCID (DSM-III-R) with 125 veterans and 50 controls to examine the prevalence of psychological and substance abuse disorders. They used a measure of test-retest reliability, the  $k$  statistic, for current and lifetime disorders. Unfortunately, the study—a brief report—did not elaborate on the details of reliability testing but did report good inter-rater agreement (100%) for 13 comparisons (ie, interviewer and observer). The test-retest  $k$  statistic was 0.61 for current disorders and 0.68 for lifetime. The study does not specify reliability by specific disorder or SCID module.

## DISCUSSION

This is the first systematic review of depression measurement in persons with SCI. Our first conclusion is that there is a dearth of psychometric data on measures used with this population. This paucity of reliability and validity data is striking in the light of the important role attributed to depression among people living with SCI. It is important to reiterate that this review focused specifically on the depression measures themselves and not on depression outcomes in this population. Only 24 studies in the last 28 years reported some type of psychometric data on depression measures; of these, 14 reported only



internal consistency. Seven studies were classified as being a validation study (level 1); this is equally surprising given the focus on depression in the SCI research literature and in clinical practice. Although the depression measures represented in the SCI literature are well validated and widely used in other populations, only 5 of the studies reviewed here examined concurrent validity with diagnostic criteria such as the DSM-III-R, and none examined convergent validity with other depression measures.

Although there is an overall paucity of psychometric data on depression measures used among people with SCI, from the evidence that is available, it seems that different measures perform equally well. For example, internal consistency across studies and measures were uniformly  $>0.70$ , concurrent validity across measures and items were generally good, and scale structures were generally multidimensional, although specific dimensions did vary between studies and measures. No single measure had evidence to place it far and above any others as the most preferable one to use. This is similar to the findings from Williams et al (27) who examined the usefulness of case-finding instruments for depression used in primary care settings. The authors examined 16 instruments (which included the CES-D, BDI, PHQ-9, and SDS) from 38 studies involving 32,000 patients, 12,900 of whom also underwent an independent diagnostic assessment. The operating characteristics of these measures were found to be quite similar. Thus selection of a particular depression measure in people with SCI cannot be made on the grounds of psychometric superiority but depends instead on feasibility, acceptability to patients, ease of administration and scoring, and the ability of the measure to serve additional purposes like monitoring response to therapy.

Despite the importance of psychometric properties of measures, the reporting of (and by extension, the evaluation of) these properties of measures across psychosocial outcomes is often woefully lacking, not only in rehabilitation (55), but in other disciplines. For example, Whittington (56) examined the ways authors fail to include adequate information about data collection in articles published in education and related journals. Among the most prevalent problems was failure to report any reliability evidence, potential problems with bias or error, evidence of content validity, qualifications or training of personnel administering measures, and when applicable, describing development of coding schemes (open-ended items) or inter-rater reliability (for open-ended items or observations). Hogan and Agnello (57) examined research reports on unpublished measurement tools in journals covering the fields of psychology, education, and sociology. Results indicated that slightly one half of nearly 700 reports had any type of validity evidence. When validity data were given, the vast majority reported correlations with other variables. In a similar study examining the reporting of reliability,

Hogan et al (58) found that reliability was reported in most reports (with the coefficient  $\alpha$  the overwhelming favorite among types of coefficients). Although reliability fared better than validity, as we found in our review here, there were still problems identified related to ambiguous designations of coefficients, reporting reliability from other studies but not the current one, and inadequate information about subscales. It seems that our current findings are, unfortunately, consistent with others studied of measurement reporting.

### **Use of Depression Measures in the SCI Literature**

Our classification of studies shows that these depression measures are typically used in tandem with other psychosocial measures (eg, quality of life, anxiety, and adjustment), with depression being one of several outcomes or predictors of interest. Fewer studies specifically focused on depression as a primary outcome and very few were exclusively concerned with the validation of a depression measure. The most frequently used measure was the BDI, closely followed by the CES-D. The other measures were used far less frequently. The PHQ-9 has been used infrequently to date, with only 2 studies reporting psychometric data; however, its inclusion in the National SCI Statistical Center Database since 2000 will likely lead its more widespread use and publications in the near future. The SCID has been used the least frequently, and this may be because of its more intensive time and training commitment.

### **Weighing the Evidence**

When examining concurrent validity with diagnostic criteria, we found that the overall sensitivity and specificity were generally good for the measures that were evaluated, although they varied slightly across studies. The BDI had the best balance of sensitivity and specificity in a sample of veterans (33); the SDS had a less favorable balance with lower specificity, but good sensitivity in a sample of traumatically injured adults (59). Both studies validated these depression severity measures against DSM-III-R criteria. None examined validity against the current DSM-IV. The examination of efficiency of IDD items against DSM-III criteria showed that most items were efficient in predicting depression though they did not all perform equally between those with paraplegia and those with tetraplegia. The study of Clay et al (53) also used negative and predictive power of each symptom, and this was a useful way to evaluate the efficiency of each of the symptoms.

The selection of an acceptable balance of sensitivity and specificity depends in part on the needs and resources of the clinician or researcher. Higher specificity may be appropriate for lower-risk populations and where resources require that the need for follow-up evaluations is minimized. For high-risk populations where identification of those with probable depression outweighs concern over the use of resources, measures and cut-offs



leaning toward higher sensitivity and lower specificity may be acceptable. Higher population prevalence in conjunction with greater risk of significant adverse consequences of untreated depression in persons with SCI may justify measures and cut-offs with greater sensitivity.

### **Scale Dimensionality of Depression Measures**

Factor analytic studies suggest that, among various samples of persons with SCI, the SDS (59), OAHMQ (60), and IDD (52) are multidimensional, whereas the CES-D is unidimensional (45). Affective dimensions were identified for the SDS and OAHMQ and to some extent for the IDD. None had a clearly delineated somatic factor. Rather, dimensions either represented a mix of psychological and somatic symptoms (SDS) (59) or were characterized by specific symptoms such as low energy (OAHMQ) (51), diminished activity (OAHMQ) (49), and slowing down (OAHMQ) (51). For the OAHMQ, cognitive dimensions also were identified separately from the affective dimension, representing helplessness, negative evaluation of life, and anhedonia (49,51). Factor analytic studies are inherently limited by being sample dependent and subject to a high level of interpretability. Nevertheless, these studies provide important insight into underlying structures of depression scales, which has particular importance for better understanding the role of somatic symptoms in depression profiles in the context of SCI.

### **Recommendations for Research**

*Testing Depression Measures Further.* Clearly, there is a need to more rigorously and widely evaluate the depression measures we are routinely using in research. Where psychometric data are available, no single measure stood out as exemplary relative to others. However, we do recommend several measures for further examination. By extrapolating from the research on patients in primary care settings (61–63) and with traumatic brain injury (64), it would seem that the PHQ-9 is a good candidate for trial as both screening and outcome measures in SCI. It is the shortest measure, is acceptable to patients, its item content exactly parallels the DSM-IV, and it has performed quite well in other populations (63–66). Although it is now part of the National SCI Statistical Center Database, it has not yet been validated for people with SCI. Its availability as part of this large database will hopefully encourage its validation in the SCI population.

Similarly, for research on the incidence and prevalence of symptoms of depression (as opposed to MDD) in the general population of persons with SCI, the CES-D has been found to be well suited for epidemiologic research with good psychometric characteristics in the general and other populations (67–71). In this review, we found it to be widely used in SCI research, second only to the BDI with consistently good internal reliability. We also

found that, although used less often than the BDI and CES-D, the IDD seems to have some promise as a valid measurement tool, in particular because it was designed to be both a measure of severity and criterion referenced to the DSM diagnostic criteria. The examination of Clay et al (53) of the efficiency of 9 of its items raised interesting questions with respect to how symptoms may function differently by level of injury and is worthy of further examination. Finally, the BDI is the most widely used of all the instruments included in this review and is worthy of further validation given its robust performance across populations.

Further validation research on these instruments for persons with SCI might yield evidence to support the use of a common measurement tool for surveillance and outcomes research. Moreover, we caution against the development of new depression measures for this population until a lack of support for existing measures is well established; instead, we encourage efforts to validate those measures we have described here and that have been shown to be valid and reliable across other clinical and general populations.

The design of these validation studies is critical because methodologic shortcomings can overestimate the accuracy of a test or measure (72). Sources of variation and bias in studies of diagnostic accuracy include demographic features, disease severity and prevalence, selection of subjects, test protocols, reference standard and verification, interpretation, and analysis (73). For example, large effects on accuracy estimation have been found in studies using cases and controls (72,74) and different reference standards for the verification of positive and negative test results (eg, gold standard for positive results and poor reference test for negative results) (74), as well as retrospective data collection (72). The effects of nonconsecutive sampling have been inconsistent, with small (74) to large (72) effects on overestimation of accuracy. We recommend that, for studies examining the diagnostic accuracy of depression measures in persons with SCI, these findings be considered carefully when designing new studies. In addition, we recommend that sample composition should take into account factors such as sex, level, and severity of injury and other injury-related characteristics with respect generalizability, bias, and variation.

Last, one of the major criticisms of studies of diagnostic accuracy is reporting standards, more specifically, the poor quality of reporting of these studies (75,76). To improve the quality of reporting, the Standards for Reporting of Diagnostic Accuracy (STARD) initiative was developed (77). We recommend that such resources are used when designing and reporting validation studies of depression measures in the SCI population.

*Symptom Profiles and Longitudinal Studies.* Studies of measures that are sensitive to change in depression symptoms in individuals with SCI also are needed.

Standard measures may lack sensitivity to change with treatment if a substantial proportion of the variance in depression scores is attributable to SCI-related symptoms, which may be the case especially soon after injury. One suggestion for untangling this challenge of accounting for somatic symptoms is to follow symptoms over time using several depression measures simultaneously to highlight the association of symptom clusters in the context of SCI. For example, if cognitive and affective symptoms change in tandem with somatic symptoms, this would provide some evidence for an association. Although the magnitude of change within these clusters may vary, a similar direction of change would support inclusion of specific somatic items in a profile of depression in the context of SCI. Alternatively, if 1 cluster of items remain stable—for example, after an intervention—whereas others change, this would provide some support for an independence of symptom clusters. To meaningfully follow symptoms across time, a measure must be sensitive to change; although studies in this review used these depression measures with time as an independent variable, sensitivity of the measure to change over time was not specifically evaluated. Another consideration in using this approach is to use the depression measures that have some evidence for being multifactorial in their structure so that clusters of symptoms can be tracked over time.

*Moving Beyond Classical Test Theory.* All of the studies reviewed here used classical test theory (CTT) methods to evaluate psychometric characteristics of the measures. Although internal consistency of measures were uniformly good to excellent across measures and studies, internal consistency, like factor analysis, is sample specific and limited in its ability to generalize across samples. The use of newer methods such as item response theory has advantages over CTT, in particular, by presuming invariance across samples, allowing for testing measurement equivalence across groups, allowing for the selection of items that provide maximum measurement precision in a specific trait range, and making it possible to examine the contribution of items individually as they are added and removed from a test. Expansion of methods to evaluate psychometric properties of depression measures is important for moving psychosocial research forward.

### **Recommendations for Clinical Practice**

*Depression Screening Programs in Rehabilitation Settings.* The rationale for implementing depression screening programs in medical settings is compelling because it is seen as an economical means of identifying people in need of psychological services based on the assumption that they will receive appropriate treatment (78); to argue against such a laudable recommendation seems unwise at first glance. However, this rationale rests on several assumptions that are not unequivocally supported by this systematic review. Most pointedly, for screening

to be efficient, screening instruments should be reliable and valid. Unfortunately, this review has shown that instruments designed specifically for depression screening—and used in such programs—have not been sufficiently validated in persons with SCI. In the 4 studies that examined concurrent validity against DSM-III-R diagnostic criteria (33,59,79), only one used a screening measure (ie, IDD) (53), and the rest used severity measures (ie, BDI, SDS, and CES-D).

The second assumption of depression screening programs is that there are resources (ie, psychologists, psychiatrists, or other mental health personnel) available to provide further evaluation and treatment for those who are identified as positively screened. The resources to interpret the positive screens can be substantial; much of this effort will be focused on screening false positives, those who are already receiving treatment, or those who do not want treatment (78). This highlights the issue of acceptable thresholds of sensitivity and specificity of depression measures. Only 4 studies in the last 28 years have examined sensitivity and specificity of depression measures against DSM criteria. Without further assessment of the concurrent validity of depression measures used in the SCI population, we are unable to determine what thresholds for these measures are indeed acceptable. We may find that, because of somatic problems or symptoms, specificity thresholds could be lower, which would entail a greater use of resources. What is the justifiable, ethical balance we should strike? There is simply not yet the empirical evidence available to reach such a consensus.

The third assumption is that screening programs assume the availability of treatment for depression and the efficacy of those interventions. The availability of and utilization of treatment for depression by persons with SCI is largely unknown. Moreover, evidence for the efficacy of depression interventions, particularly pharmacologic, is woefully lacking for the SCI population (80), limited in part by the paucity of evidence we have discussed in this review. Finally, the acceptability of screening programs to key stakeholders, such as patients, physicians, and support staff, is largely unknown (81).

Given these caveats and the results of this review, the state of science for depression measurement in SCI makes a recommendation for the implementation of universal depression screening programs premature despite the importance of identifying and treating depression in this population. Furthermore, it illustrates how inextricably linked measurement is to clinical practice and the importance of developing psychometrics in this area.

*Screening for Depression in Clinical Settings.* Although a recommendation for comprehensive depression screening programs in rehabilitation settings where persons with SCI may receive care is by our estimation premature, screening for depression remains important in clinical settings. Recommendations for the use of

standardized criteria in clinical practice are not new (25). The results of this review do not support the recommendation of one measure over another in clinical practice. Individual clinicians will have to base their decisions on the particular needs of their patient, the scope of their practice and resources available. We do advocate for the integration of standardized measures into clinical assessments, particularly where follow-up may be handled exclusively by medical personnel, because often individuals are only seen by their physicians on a regular basis rather than clinical psychologists or other mental health personnel.

Unfortunately, we know little about how and the degree to which depression measures are actually used in clinical practice or how clinicians select the measures they use. Knowing how tools are used is as important as knowing about their reliability and validity. Furthermore, this review can only address the use of depression measures in research studies in the published literature—their use in clinical practice is far more difficult to assess. We do recommend that, when measures are considered by clinicians, those with some support for reliability and/or validity are considered over those with no available data. Furthermore, established measures are preferred over the use of nonstandardized or untested clusters of questions.

### Limitations of this Review

As we have noted earlier, because there are no standards for conducting a systematic review of measurement tools, our protocol evolved over the course of the review. Although it is ideal that a review can be replicated, this is much more easily achieved with more narrow parameters such as targeting the selection of only randomized controlled trials about a specific intervention. Here, replication is challenged by inherent subjectivity in the selection of studies to be included in the review. For example, our decision to include any psychometric data—which was largely represented by internal consistency as we have reported—may be too wide a net for other investigators. The a priori selection of specific depression measures may have been an unnecessary step. This review is also limited to the English, peer-reviewed published literature. Finally, we did not judge the quality of the studies that were included in this review.

The initial classification into levels of the studies was meant only as a way to organize the 377 studies that were to be evaluated. Also, only the abstracts were used to classify studies and, because the focus was on extracting psychometric data, checking each study's classification was not addressed once the full article was read. In retrospect, it was not an ideal way to organize the information, although it did not impede our ability to identify psychometric data. Instead its value was found to be in summarizing the ways in which depression measures were used in the SCI literature. The checking and reclassification when necessary of the studies after

the data were extracted was not part of the original plan for this review and is unlikely to be exactly replicable, because a number of studies did not clearly explicate whether depression was of primary or secondary interest as an outcome, and therefore, a certain degree of subjective judgment from a single author was used to reclassify studies.

### CONCLUSIONS

The results of this first systematic review of depression measures in SCI and the recommendations based on these findings show that there is much work to be done. It is important to establish the credentials of a uniform measure of depression so that outcomes of clinical research can easily be compared across studies. Although we may not find dramatic differences in the reliability, validity, and clinical efficiency of the depression measures reviewed here when used in the SCI population, there is still much work to be done to establish their psychometric soundness in this population. If continued unchecked, this gap in our knowledge will impede our ability to validly and reliably identify individuals with SCI who are depressed, achieve reliable estimates of incidence and prevalence of depression in this population, and understand the role of trans-diagnostic symptoms. Moreover, our ability to target interventions on the most problematic symptoms, examine the effectiveness of our interventions, consider and evaluate the implementation of depression screening programs, and effectively use measurement tools in clinical practice will be limited. Simply said, the value of continued work evaluating depression measures for the SCI population cannot be understated.

### ACKNOWLEDGMENTS

The authors thank Grace Wang, MPH, Elizabeth Webber, MS, and Melanie Feinberg, MIMS, for work on this review.

### REFERENCES

1. Elliott TR, Frank RG. Depression following spinal cord injury. *Arch Phys Med Rehabil*. 1996;77(8):816–823.
2. Bombardier CH, Richards JS, Krause JS, Tulsy D, Tate DG. Symptoms of major depression in people with spinal cord injury: implications for screening. *Arch Phys Med Rehabil*. 2004;85(11):1749–1756.
3. Malec J, Neimeyer R. Psychologic prediction of duration of inpatient spinal cord injury rehabilitation and performance of self-care. *Arch Phys Med Rehabil*. 1983;64(8):359–363.
4. Herrick S, Elliott T, Crow F. Social support and the prediction of health complications among persons with spinal cord injuries. *Rehabil Psychol*. 1994;39(4):231–250.
5. Elliott T, Shewchuck R. Social support and leisure activities following severe physical disability: testing the mediating effects of depression. *Basic Appl Soc Psych*. 1995;16:471–487.
6. Fuhrer JM, Rintala DH, Hart KA. Depressive symptomatology in persons with spinal cord injury who reside in the community. *Arch Phys Med Rehabil*. 1993;74:255–260.

7. MacDonald M, Nielson W, Cameron M. Depression and activity patterns of spinal cord injured persons living in the community. *Arch Phys Med Rehabil.* 1987;68(6):339–343.
8. Schulz R, Decker S. Long-term adjustment to physical disability: the role of social support, perceived control, and self-blame. *J Pers Soc Psychol.* 1985;48(5):1162–1172.
9. Consortium for Spinal Cord Medicine. Clinical Practice Guideline: depression following spinal cord injury: a clinical practice guideline for primary care physicians. *J Spinal Cord Med.* 2001;24(suppl 1):S40–S101.
10. Munoz R, Le H, Ippen C. We should screen for major depression. *Appl Prev Psychol.* 2000;9(2):123–133.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association; 1994.
12. Zimmerman M, Chelminski I, McGlinchey JB, Young D. Diagnosing major depressive disorder VI: performance of an objective test as a diagnostic criterion. *J Nerv Ment Dis.* 2006;194(8):565–569.
13. Beck AT, Ward C, Mendelson M. Beck Depression Inventory (BDI). *Arch Gen Psychiatry.* 1961;4:561–571.
14. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56–62.
15. Zung W. A self-rating depression scale. *Arch Gen Psychiatry.* 1965;12:63–70.
16. Spitzer R, Kroenke K, Williams J. Validity and utility of a self-report version of PRIME-MD: the PHQ Primary Care Study. *JAMA.* 1999;282:1737–1744.
17. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I/NP), Research Version*. New York: Biometrics Research, New York State Psychiatric Institute; 2002.
18. Stewart DE. Physical symptoms of depression: unmet needs in special populations. *J Clin Psychiatry.* 2003;64:12–16.
19. Williams JW, Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA.* 2002;287(9):1160–1170.
20. Greden JF. Physical symptoms of depression: unmet needs. *J Clin Psychiatry.* 2003;64:5–11.
21. Kroenke K. The interface between physical and psychological symptoms. *Prim Care Companion J Clin Psychiatry.* 2003;5(suppl 7):11–18.
22. Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. *N Engl J Med.* 1999;341(18):1329–1335.
23. Kroenke K, Spitzer R, Williams J, et al. Physical symptoms in primary care. Predictors of psychiatric disorders and functional impairment. *Arch Fam Med.* 1994;3(9):774–779.
24. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. *Arch Intern Med.* 2003;163(20):2433–2445.
25. Frank R, Elliott T, Corcoran J, Wonderlich S. Depression after spinal cord injury: is it necessary? *Clin Psychol Rev.* 1987;7(6):611–630.
26. Jacob KS, Zachariah K, Bhattacharji S. Depression in individuals with spinal cord injury: methodological issues. *Paraplegia.* 1995;33(7):377–380.
27. Williams J, Pignone M, Ramirez G, Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry.* 2002;24(4):225–237.
28. Haywood KL, Garratt AM, Fitzpatrick R. Older people specific health status and quality of life: a structured review of self-assessed instruments. *J Eval Clin Pract.* 2005;11(4):315–327.
29. Bot SDM, Terwee CB, van der Windt D, Bouter LM, Dekker J, de Vet HCW. Clinometric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis.* 2004;63(4):335–341.
30. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193–205.
31. Radnitz CL, Broderick CP, Perez-Strumolo L, et al. The prevalence of psychiatric disorders in veterans with spinal cord injury: a controlled comparison. *J Nerv Ment Dis.* 1996;184(7):431–433.
32. Judd FK, Stone J, Webber JE. Depression following spinal cord injury: a prospective in-patient study. *Br J Psychiatry.* 1989;154:668–671.
33. Radnitz CL, McGrath RE, Tirch DD, et al. Use of the Beck Depression Inventory in veterans with spinal cord injury. *Rehabil Psychol.* 1997;42(2):93–101.
34. Ferington FE. Personal control and coping effectiveness in spinal cord injured persons. *Res Nurs Health.* 1986;9(3):257–265.
35. Tate DG, Forchheimer M, Maynard F, Davidoff G, Dijkers M. Comparing two measures of depression in spinal cord injury. *Rehabil Psychol.* 1993;38(1):53–61.
36. Latimer AE, Ginis KAM, Hicks AL. Buffering the effects of stress on well-being among individuals with spinal cord injury: a potential role for exercise. *Ther Recreation J.* 2005;39(2):131–138.
37. Latimer AE, Ginis KAM, Hicks AL, McCartney N. An examination of the mechanisms of exercise-induced change in psychological well-being among people with spinal cord injury. *J Rehabil Res Dev.* 2004;41(5):643–651.
38. Hicks AL, Martin KA, Ditor DS, et al. Long-term exercise training in persons with spinal cord injury: effects on strength, arm ergometry performance and psychological well-being. *Spinal Cord.* 2003;41(1):34–43.
39. Ginis KAM, Latimer AE, McKechnie K, et al. Using exercise to enhance subjective well-being among people with spinal cord injury: the mediating influences of stress and pain. *Rehabil Psychol.* 2003;48(3):157–164.
40. Decker SD, Schulz R. Correlates of life satisfaction and depression in middle-aged and elderly spinal cord-injured persons. *Am J Occup Ther.* 1985;39(11):740–745.
41. Coyle CP, Shank JW, Kinney W, Hutchins DA. Psychosocial functioning and changes in leisure lifestyle among individuals with chronic secondary health problems related to spinal cord injury. *Ther Recreation J.* 1993;27(4):239–252.
42. Dorsett P, Geraghty T. Depression and adjustment after spinal cord injury: a three-year longitudinal study. *Top Spinal Cord Inj Rehabil.* 2004;9(4):43–56.
43. Crisp R. Locus of control as a predictor of adjustment to spinal cord injury. *Austr Dis Rev.* 1984;1(2):53–57.
44. Rintala DH, Robinson-Whelen S, Matamoros R. Subjective stress in male veterans with spinal cord injury. *J Rehabil Res Dev.* 2005;42(3):291–304.
45. McColl MA, Skinner HA. Measuring psychological outcomes following rehabilitation. *Can J Public Health.* 1992;83(suppl 2):S12–S18.
46. Shnek ZM, Foley FW, LaRocca NG, et al. Helplessness, self-efficacy, cognitive distortions, and depression in multiple



- sclerosis and spinal cord injury. *Ann Behav Med*. 1997; 19(3):287–294.
47. Kuptniratsaikul V, Pekuman P. The Study of the Center for Epidemiologic Studies Depression Scale in Thai People. *Siriraj Hospital Gaz*. 1997;49:442–448.
  48. Kemp BJ, Kahan JS, Krause JS, Adkins RH, Nava G. Treatment of major depression in individuals with spinal cord injury. *J Spinal Cord Med*. 2004;27(1):22–28.
  49. Krause J, Kemp BJ, Coker J. Depression after spinal cord injury: relation to gender, ethnicity, aging, and socioeconomic indicators. *Arch Phys Med Rehabil*. 2000;81(8):1099–1109.
  50. Krause JS, Coker J, Charlifue S, Whiteneck GG. Depression and subjective well-being among 97 American Indians with spinal cord injury: a descriptive study. *Rehabil Psychol*. 1999;44(4):354–372.
  51. Kemp B, Krause J, Adkins R. Depression among African Americans, Latinos, and Caucasians with Spinal cord injury: an exploration study. *Rehabil Psychol*. 1999;44:235–247.
  52. Frank R, Chaney J, Clay D. Dysphoria: a major symptom factor in persons with disability or chronic illness. *Psychiatry Res*. 1992;43:231–241.
  53. Clay DL, Hagglund KJ, Frank RG, Elliott TR, Chaney JM. Enhancing the accuracy of depression diagnosis in patients with spinal-cord injury using Bayesian-analysis. *Rehabil Psychol*. 1995;40(3):171–180.
  54. Kalpakjian C, Albright K. An examination of depression through the lens of spinal cord injury: comparative prevalence rates and severity in women and men. *Womens Health Issues*. 2006;16:380–388.
  55. Dijkers M, Kropp GC, Esper RM, Yavuzer G, Cullen N, Bakdalieh Y. Reporting on reliability and validity of outcome measures in medical rehabilitation research. *Disabil Rehabil*. 2002;24(16):819–827.
  56. Whittington D. How well do researchers report their measures? An evaluation of measurement in published educational research. *Educ Psychol Meas*. 1998;58(1):21–37.
  57. Hogan TP, Agnello J. An empirical study of reporting practices concerning measurement validity. *Educ Psychol Meas*. 2004;64(5):802–812.
  58. Hogan TP, Benjamin A, Brezinski KL. Reliability methods: a note on the frequency of use of various types. *Educ Psychol Meas*. 2000;60(4):523–531.
  59. Tate D, Forchheimer M, Maynard F, Davidoff G, Dijkers M. Comparing two measures of depression in spinal cord injury. *Rehabil Psychol*. 1993;38:53–61.
  60. Kemp BJ, Adams BM. The Older Adult Health and Mood Questionnaire: a measure of geriatric depressive disorder. *J Geriatr Psychiatry Neurol*. 1995;8(3):162–167.
  61. Lowe B, Schenkel I, Carney-Doebbeling C, Gobel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics*. 2006;47(1):62–67.
  62. Lowe B, Kroenke K, Grafe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res*. 2005;58(2):163–171.
  63. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2006;28(1):71–77.
  64. Fann JR, Bombardier CH, Dikmer S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil*. 2005;20(6):501–511.
  65. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord*. 2004;81(1):61–66.
  66. Williams LS, Brizendine EJ, Plue L, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 2005;36(3):635–638.
  67. McCauley SR, Pedroza C, Brown SA, et al. Confirmatory factor structure of the Center for Epidemiologic Studies-Depression scale (CES-D) in mild-to-moderate traumatic brain injury. *Brain Inj*. 2006;20(5):519–527.
  68. Rhee SH, Petroski GF, Parker JC, et al. A confirmatory factor analysis of the Center for Epidemiologic Studies Depression Scale in rheumatoid arthritis patients: additional evidence for a four-factor model. *Arthritis Care Res*. 1999;12(6):392–400.
  69. Knight RG, Williams S, McGee R, Olan S. Psychometric properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behav Res Ther*. 1997;35(4):373–380.
  70. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. *Am J Epidemiol*. 1977; 106(3):203–214.
  71. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging*. 1997;12(2):277–287.
  72. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469–476.
  73. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140(3):189–202.
  74. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061–1066.
  75. Bossuyt PMM. The quality of reporting in diagnostic test research: getting better, still not optimal. *Clin Chem*. 2004; 50(3):465–466.
  76. Rennie D. Improving reports of studies of diagnostic tests: the STARD initiative. *JAMA*. 2003;289(1):89–90.
  77. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med*. 2003;138(1):40–44.
  78. Coyne JC, Thompson R, Palmer SC, Kagee A, Maunsell E. Should we screen for depression? Caveats and potential pitfalls. *Appl Prevent Psychol*. 2000;9:101–121.
  79. Kuptniratsaikul V. Epidemiology of spinal cord injuries: a study in the Spinal Unit, Siriraj Hospital, Thailand, 1997–2000. *J Med Assoc Thai*. 2003;86(12):1116–1121.
  80. Elliott TR, Kennedy P. Treatment of depression following spinal cord injury: an evidence-based review. *Rehabil Psychol*. 2004;49(2):134–139.
  81. Palmer SC, Coyne JC. Screening for depression in medical care: pitfalls, alternatives, and revised priorities. *J Psychosom Res*. 2003;54(4):279–287.
  82. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measure*. 1977;1(3):385–401.



83. Zimmerman M, Coryell W, Coryell C, Wilson S. A self-report scale to diagnose major depressive disorder. *Arch Gen Psychiatry*. 1986;43(11):1076–1081.
84. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. 1999;282(18):1737–1744.
85. First M, Gibbon M, Spitzer R, Williams J. *User's Guide for the Structured Clinical Interview for DSM IV Axis I Disorders*. New York: Biometrics Research Department, New York State Psychiatric Institute; 1996.
86. Schulz R, Decker S. Long-term adjustment to physical disability: the role of social support, perceived control, and self-blame. *J Pers Soc Psychol*. 1985;48(5):1162–1172.

## Appendix: Databases and Search Terms

Database	Search Terms
PubMed	Spinal cord injury AND (depression OR (CESD OR Zung OR SCID OR “Older Adult Mood and Health” OR DSM OR BDI OR PHQ-9)) <i>In PubMed, terms are searched as both keywords and subject headings simultaneously.</i>
CINAHL	Spinal cord injuries AND (depression OR (Center for Epidemiological Studies Scale OR DSM OR Beck Depression Inventory OR PHQ-9* OR “Older Adult Mood and Health”* OR SCID* OR Zung*)) <i>Terms with an asterisk did not have associated CINAHL subject headings and were searched as keywords; other terms were searched as CINAHL subject headings.</i>
PsycINFO	Spinal cord injuries AND (major depression OR (Diagnostic and Statistical Manual OR Zungs Self Rating Scale OR Beck Depression Inventory OR Structured Clinical Interview OR PHQ-9* OR “Older Adult Mood and Health”* OR CESD*)) <i>Terms with an asterisk did not have associated PsycINFO subject headings and were searched as keywords; other terms were searched as PsycINFO subject headings.</i>
ProQuest Health and Medical Complete Library	“Spinal cord injury” AND depression  <i>Terms searched as keywords, not ProQuest subject headings. Adding terms for scales did not increase relevant results for this database.</i>
Web of Science	“Spinal cord injury” AND (“major depression” OR “depressive disorder” OR (CESD OR Zung OR SCID OR “Older Adult Mood and Health” OR DSM OR BDI OR PHQ-9)) <i>Terms searched as keywords.</i>
Google Scholar	Spinal-cord-injury CESD OR BDI OR PHQ-9 OR DSM OR Zung OR “older OR adult OR mood OR and OR health” OR SCID “Major depression” <i>Search limited to Medicine, Pharmacology, and Veterinary Science</i>